5   receiving, from the user, data representative of one or more selected categorical labels;

6   and

7   labelling the document within the collection with the one or more selected categorical

8   labels.

1   12.   The method of claim 11 further comprising the step of deriving a plurality of

2   categorization shortcuts from the plurality of most likely categorical labels, wherein the displaying

3   step comprises the step of displaying, to the user, the plurality of categorization shortcuts.

13.   The method of claim 11 wherein the classifying step comprises the step of classifying,

2   upon receipt, the document to obtain the plurality of most likely categorical labels.

1   14.   The method of claim 12 wherein the deriving step comprises the step of deriving, upon

2   receipt of the document, categorization shortcuts from the plurality of most likely categorical

3   labels.

1   15.   The method of claim 12 wherein the deriving step comprises the step of labelling display

2   buttons with the plurality of most likely categorical labels, and the displaying step comprises the

3   step of displaying the labelled display buttons with the document.

16.   The method of claim 12 wherein the deriving step comprises the step of creating an

2   ordered set of the plurality of most likely categorical labels, and the displaying step comprises the

3   step of displaying with the document the ordered set prepended to a standard ordering of other

4   categorical labels.

1   17.   The method of claim 11 wherein the classifying step occurs substantially simultaneously

2   with the displaying step.

1   18.   The method of claim 11 wherein the classifying step comprises the step of classifying,

2   upon invocation by the user, the document to obtain the plurality of most likely categorical labels.

1   19.   The method of claim 18 wherein the invocation comprises a selection by the user of a

2   classify button.

1   20.   The method of claim 11 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3   categorical labels.

1   21.   The method of claim 11 further comprising the step of displaying a standard list of all

2   categorical labels, wherein the receiving step comprises the step of receiving, from the user, data

3   representative of one or more selected categorical labels from either the plurality of displayed

4   categorization shortcuts or the standard list.

1   22.   The method of claim 11 wherein the classifying step is performed by a classifier and

2   further comprising the step of incrementally re-training the classifier to adapt to modifications of

3   the collection.

SUB B4

23.   The method of claim 22 wherein the re-training step comprises the step of re-training the

2   classifier in response to the labelling step.

1   24.   The method of claim 22 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3   categorical labels and the re-training step comprises the steps of:

A1

4   receiving, from the user, addition data representative of an addition of a document into a

5   tofolder; and

6   re-training the classifier in response to the addition data.

1   25.   The method of claim 24 wherein the re-training step comprises the step of assigning, in the

2   classifier, the added document to the tofolder.

1   26.   The method of claim 25 further comprising the step of identifying excluded folders to be

2   excluded from re-training and wherein the re-training step comprises the step of assigning, in the

3   classifier, the added document when the tofolder is not one of the identified excluded folders.

SUB B5

27.   The method of claim 22 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3 categorical labels and the re-training step comprises the steps of:

4   receiving, from the user, deletion data representative of a removal of a document from a

5 fromfolder; and

6   re-training the classifier in response to the deletion data.

1 28.   The method of claim 27 wherein the re-training step comprises the step of unassigning, in

2 the classifier, the removed document from the fromfolder in which it was categorized.

1 29.   The method of claim 28 further comprising the step of identifying excluded folders to be

2 excluded from re-training and wherein the re-training step comprises the step of unassigning, in

3 the classifier, the removed document when the fromfolder is not one of the identified excluded

4 folders.

SUB 186   30.   The method of claim 22 wherein the labelling step comprises the step of storing the

2 document in folders or locations of the collection corresponding to the one or more selected

3 categorical labels and the re-training step comprises the steps of:

4   receiving, from the user, move data representative of a movement of a document from a

5 source folder to a destination folder; and

6   re-training the classifier in response to the move data.

1 31.   The method of claim 30 wherein the re-training step comprises the steps of:

2   unassigning, in the classifier, the moved document from the source folder in which it was

5

3   categorized; and

4   assigning, in the classifier, the moved document to the destination folder..

1   32.   The method of claim 31 further comprising the step of identifying excluded folders to be

2   excluded from re-training and wherein the re-training step comprises the steps of:

3   unassigning, in the classifier, the moved document when the source folder is not one of the

4   identified excluded folders; and

5   assigning, in the classifier, the moved document when the destination folder is not one of

6   the identified excluded folders.

4/   1   33.   The method of claim 22 wherein the re-training step occurs instantly after a collection

SUB 27/   modification.

1   34.   The method of claim 22 wherein the re-training step occurs a fixed amount of time after a

2   last re-training or an initial training from scratch.

1   35.   The method of claim 22 wherein the re-training step occurs when a threshold number of

2   documents have been added, deleted or moved in the collection.

1   36.   The method of claim 22 wherein the re-training step occurs when an idle state is reached.

1   37.   The method of claim 20, wherein the classifying step comprises the steps of:

6

2   tokenizing the document into different tokens;

3   tallying a number of occurrences of each token in the document;

4   computing, for each folder, a token weight of each token;

5   comparing, for each token, the number of occurrences and the token weights;

6   creating a similarity score in response to the comparing step; and

7   identifying a subset of folders for which the similarity score is highest.

1   38.   The method of claim 37 further comprising the step of removing, from the identified

2   subset, all folders for which the similarity score is lower than a default or specified threshold.

Al

1   39.   The method of claim 37, wherein the computing step comprises the step of computing the

2   token counts of each token in each of the folders.

1   40.   The method of claim 37 wherein the tokenizing step comprises the steps of:

2   separately tokenizing different portions of the document; and

3   labelling the tokens according to the different portions.

1   41.   The method of claim 25 wherein the classifying step comprises the steps of:

2   tokenizing the document into different tokens;

3   tallying a number of occurrences of each token in the document;

4   retrieving, for each folder, a tokencount of each token;

5   computing, for each folder, a token weight of each token;

6   comparing, for each token, the number of occurrences and the token weights;

7   creating a similarity score in response to the comparing step; and

8   identifying a subset of folders for which the similarity score is highest, and

9   wherein the assigning step comprises the step of adding the number of occurrences of each token

10   to the tokencount of the tofolder.

1   42.   The method of claim 28 wherein the classifying step comprises the steps of:

2   tokenizing the document into different tokens;

3   tallying a number of occurrences of each token in the document;

4   retrieving, for each folder, a tokencount of each token;

5   computing, for each folder, a token weight of each token;

6   comparing, for each token, the number of occurrences and the token weights;

7   creating a similarity score in response to the comparing step; and

8   identifying a subset of folders for which the similarity score is highest, and

9   wherein the unassigning step comprises the step of subtracting the number of occurrences of each

10   token from the tokencount of the fromfolder.

1   43.   The method of claim 31 wherein the classifying step comprises the steps of:

2   tokenizing the document into different tokens;

3   tallying a number of occurrences of each token in the document;

4   retrieving, for each folder, a tokencount of each token;

5   computing, for each folder, a token weight of each token;

6   comparing, for each token, the number of occurrences and the token weights;

7   creating a similarity score in response to the comparing step; and

8   identifying a subset of folders for which the similarity score is highest, and

9   wherein the unassigning step comprises the step of subtracting the number of occurrences of each

10   token from the tokencount of the source folder, and the assigning step comprises the step of

11   adding the number of occurrences of each token to the tokencount of the destination folder.

1   44.   The method of claim 11, further comprising the step of training the classifier from scratch

2   with a pre-existing collection of categorized documents.

1   45.   The method of claim 44 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3   categorical labels and the training step comprises the step of assigning, in the classifier, each of

4   the pre-existing documents to a folder in which it is categorized.

1   46.   The method of claim 45 wherein the classifying step comprises the steps of:

2   tokenizing the document into different tokens;

3   tallying a number of occurrences of each token in the document;

4   retrieving, for each folder, a tokencount of each token;

5   computing, for each folder, a token weight of each token;

6   comparing, for each token, the number of occurrences and the token weights;

7   creating a similarity score in response to the comparing step; and

8   identifying a subset of folders for which the similarity score is highest, and

9   wherein the assigning step comprises the step of adding the number of occurrences of each token

10   to the tokencount of the tofolder.

1   47.   The method of claim 45 further comprising the step of identifying excluded folders to be

2   excluded from training and wherein the training step comprises the step of assigning, in the

3   classifier, each of the pre-existing documents, except those in the identified excluded folders.

1   48.   The method of claim 11 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3   categorical labels and the re-training step comprises the steps of:

4   determining a time of a last step of re-training; and

5   retraining the classifier on each folder which was modified after the determined time.

1   49.   The method of claim 22 wherein the labelling step comprises the step of storing the

2   document in folders or locations of the collection corresponding to the one or more selected

3   categorical labels, the method further comprising the step of training the classifier from scratch

4   with a pre-existing collection of categorized documents, wherein the re-training step comprises

5   the steps of:

6   determining a time of the step of training or a last step of re-training; and

7   retraining the classifier on each folder which was modified after the determined time.

1   50.   The method of claim 11 wherein the classifying step uses the TF-IDF principle.

1   51.   The method of claim 11 wherein the electronic document is an e-mail message.

1   52.   The method of claim 11 wherein the electronic document is a web page and the collection

2   is a collection of bookmarks.

1   53.   The method of claim 41 wherein the electronic document is a web page and the collection

2   is a collection of bookmarks, the method further comprising the step of storing, for each web

3   page, a pagetokencount matching the tallied number of occurrences of each token.

1   54.   The method of claim 42 wherein the electronic document is a web page and the collection

2   is a collection of bookmarks, the method further comprising the step of storing, for each web

3   page, a pagetokencount matching the tallied number of occurrences of each token, wherein the

4   unassigning step comprises the step of subtracting the pagetokencount from the tokencount of the

5   fromfolder.

1   55.   The method of claim 43 wherein the electronic document is a web page and the collection

2   is a collection of bookmarks, the method further comprising the step of storing, for each web

3   page, a pagetokencount matching the tallied number of occurrences of each token, wherein the

4   unassigning step comprises the step of subtracting the pagetokencount from the tokencount of the

5   fromfolder.

11

1   56.   The method of claim 11 wherein the electronic document is a multimedia document.

1   57.   The method of claim 56 wherein the multimedia document is an image file, a video file or

2   an audio file.

1   58.   The method of claim 56 wherein the multimedia document combines any combination of

2   text, an image file, a video file and an audio file.

A

1   59.   The method of claim 57 wherein the multimedia document further includes text.

1   60.   The method of claim 11 wherein the electronic document comprises data sets that are not

2   viewable in their entirety, but can be categorized in response to some presentation to the user.

1   61.   A program storage device, readable by a machine, tangibly embodying a program of

2   instructions executable by the machine to perform method steps for assisting a user with the task

3   of categorizing an electronic document into a collection according to the method steps of claim

4   11.--.